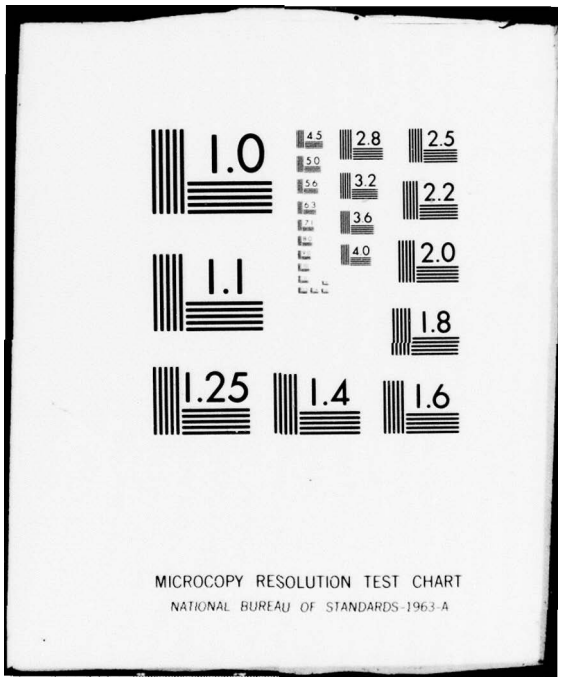
AD-A062 750 PRINCETON UNIV N J DEPT OF STATISTICS F/G 12/2
ASPECTS OF OPTIMAL DESIGN IN DYNAMIC SYSTEMS.(U)
SEP 78 D M TITTERINGTON N00014-75-C-0453
UNCLASSIFIED TR-141-SER-2 NL

END
DATE FILMED
3-79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL II

(6)

SC

AD A062750

DDC FILE COPY

(9) ASPECTS OF OPTIMAL DESIGN IN DYNAMIC SYSTEMS.

by

D. M. Titterington*
Department of Statistics
University of Glasgow
Glasgow, Scotland

(14) TR-141-SER-2

(9) Technical Report, No. 141, Series 2
Department of Statistics
Princeton University
(11) September 1978 (12) 35p.

(15) Research supported in part by a contract with the Office of Naval Research, No. N00014-75-C-0453, awarded to the Department of Statistics, Princeton University.

*This report was prepared by Dr. (10) D. M. Titterington during a brief visit to Princeton University, Princeton, New Jersey.

406 873

LB

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TECHNICAL REPORT	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER T.R. # 141, Series 2
4. TITLE (and Subtitle) ASPECTS OF OPTIMAL DESIGN IN DYNAMIC SYSTEMS		5. TYPE OF REPORT & PERIOD COVERED Technical paper 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) D. M. TITTERINGTON		8. CONTRACT OR GRANT NUMBER(s) ONR N000-14-75-C-0453
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Princeton University Princeton, N.J. 08540		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 471) Arlington, Virginia 22217		12. REPORT DATE September 1978 13. NUMBER OF PAGES 33
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. [illegible]

18. SUPPLEMENTARY NOTES

This report was prepared by Dr. D. M. Titterington during a brief visit to Princeton University, Princeton, New Jersey. Dr. Titterington is associated with the University of Glasgow, Glasgow, Scotland.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

dynamic systems; input signal synthesis; optimal experimental design; parameter estimation.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper attempts to provide an introduction for statisticians to the version of optimal experimental design theory for parameter estimation in regression models that is appropriate to dynamic systems. The paper consists of three main parts: first, a glossary of some terminology in control engineering and an introduction to the main aspects of dynamic systems; second, a summary of the principal results and patterns in optimal experimental design theory; and third, the ways in which the latter carry over to dynamic models. These application are split roughly into those in which sampling items are selected.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

This paper attempts to provide an introduction for statisticians to the version of optimal experimental design theory for parameter estimation in regression models that is appropriate to dynamic systems. The paper consists of three main parts: first, a glossary of some terminology in control engineering and an introduction to the main aspects of dynamic systems; second, a summary of the principal results and patterns in optimal experimental design theory; and third, the ways in which the latter carry over to dynamic models. These applications are split roughly into those involving choice of input functions and those in which sampling times are selected.

KEY WORDS: dynamic systems; input signal synthesis; optimal experimental design; parameter estimation.

1. INTRODUCTION

The main objective of this work is to bring to a statistical readership the recent activity among control engineers in the field of experimental design. Principally, we mean to catalogue work that parallels so-called optimal experimental design in which important references are Kiefer (1959, 1974), Kiefer and Wolfowitz (1960), Fedorov (1972), Whittle (1973), and Silvey (1974). It is to be hoped that statisticians will become more familiar with and more interested in dynamic problems, a hope that has been expressed before in Wishart (1969), Young (1975), Wynn (1974), and Harrison and Stevens (1976). In the engineering literature there are useful surveys by Mehra (1974b) and Goodwin and Payne (1977), but it is hoped that a "translation" might be helpful to a statistical audience.

The dictionary for the translation is provided in Section 2. In Section 3 the main features of "static" optimal design theory are laid out and in Sections 4 and 5 generalizations of these to dynamic systems are described. Section 6 contains a brief conclusion.

2. SOME TERMINOLOGY

A major discouragement to statisticians who approach the engineering literature is the "wealth" of jargon. There are both new concepts and alternative terms for familiar ideas. Many of these are discussed at length in the survey by Wishart (1969) of the deterministic optimal control problem, and here we give but a brief introduction to the new language, with special regard to the problems related to experimental designs.

Since "time" is an essential feature of dynamic systems, we will

be concerned with stochastic processes, which may be described in discrete- or continuous-time and which may be uni- or multi-variate, stationary or nonstationary.

The system itself, which the statistician would be more likely to call the model, generally involves processes of three types: inputs, outputs, and noise. (We shall see later that a fourth category, the state, is often used, but it arises less directly and we delay its description for the time being.) The inputs, or controls, are generally open to choice, the outputs may be observed by the experimenter, and the noise is random disturbance, which may be observation error or a contribution to the dynamic evolution of the process.

A further component of the system is a set of parameters, conceptually familiar to the statistician.

As an exercise in the terminology, let us consider the following simple model.

$$y(t) - a_1(t)y(t-1) - a_2(t)y(t-2) = b_1(t)u(t) + e(t),\ t = 1,2,\ldots \quad (1)$$

with some initial conditions such as $y(0) = y(-1) = 0$. $\{y(t)\}$ are the outputs, $\{u(t)\}$ are the inputs, $\{e(t)\}$ the noise and the parameters are $\{a_1(t), a_2(t), b_1(t)\}$, along with the statistical description of the noise process, which almost always has zero means. In most discrete-time problems the noise is assumed to be normally distributed (Gaussian to the engineers) and if the $\{e(t)\}$ are uncorrelated and identically distributed, the noise is said to be white because of its consequently flat spectral density. Often

$$a_1(t) = a_1 , \quad \text{for all } t,$$

and similarly for the $\{a_2(t), b_1(t)\}$. The system is then called

time-invariant, as opposed to time-varying and (1) becomes

$$y(t) - a_1 y(t-1) - a_2 y(t-2) = b_1 u(t) + e(t) , \quad t = 1,2,\ldots . \qquad (2)$$

Let us suppose for the moment that the input and output processes are scalars. Then the control engineers would describe (2) as the input-output representation of a linear, time-invariant, discrete-time, stochastic, single-input-single-output system. The antonyms of all the adjectives are obvious. By "linear" is meant linearity in the processes, not the parameters, although, apart from the parameters in the noise process, we do have this sort of linearity as well. If the $\{e(t)\}$ are normally identically and independently distributed with zero means, we might augment the description by adding that the system is "driven by white Gaussian noise."

Of course, for the above example the familiar time-series language of Box and Jenkins (1976) is also used, and the concept of stationarity is also of concern to the engineers.

The recursive nature of (2) leads to the possibility of constructing a generating function version. Thus if we denote by

$$Y(z) = \sum_{i=0} z^i y(i) ,$$

the z-transform of the output process, and so on, and if we take $u(0) = 0$, (2) can be written

$$A(z)\, Y(z) = B(z)U(z) + E(z) ,$$

where

$$A(z) = a_0 - a_1 z - a_2 z^2$$

and

$$B(z) = b_1 .$$

Going a stage further, we have, if $A(z)^{-1}$ exists,

$$Y(z) = H_1(z)U(z) + H_2(z)E(z) ,$$

where $H_1(z) = A(z)^{-1}B(z)$ and $H_2(z) = A(z)^{-1}$ are called the transfer functions from the input and the noise, respectively, to the output. For more discussion of this formulation, see Wishart (1969) and Cadzow (1973).

A further important concept is that of state-space models, which revolutionized control theory methodology. The principal objective, in discrete-time systems, is to write the model description as a set of first-order recursions on the so-called state variable(s), coupled with an equation relating the observation or output at time t with the state variable(s) and input variable(s) at time t. We should therefore have, for a linear time-invariant system, a model of the form

$$\left.\begin{aligned} \underline{x}(t+1) &= G\underline{x}(t) + Hu(t) + Fe(t) \\ y(t) &= B\underline{x}(t) + Cu(t) + Dn(t) , \quad t = 1,2,... \end{aligned}\right\} \qquad (3)$$

where $\underline{x}(t)$ denotes the vector of state variables at time t and $\{e(t)\}$ and $\{n(t)\}$ are noise processes.

Systems can have both an input-output and a state-space representation. For (2), if we define two state variables in

$$x_1(t) = y(t-1)$$
$$x_2(t) = y(t-2) ,$$

then we can replace (2) by (3) with

$$G = \begin{bmatrix} a_1 & a_2 \\ 1 & 0 \end{bmatrix}, \; H = \begin{bmatrix} b_1 \\ 0 \end{bmatrix}, \; F = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \; B^T = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \; C = (b_1),$$

$D = (1)$ and $n(t) = e(t)$, $t = 1,2,...,$ along with the initial

condition $\underline{x}(1) = \underline{0}$.

The state-space representation is important for various reasons. It is easy to develop the model sequentially in the time-domain. Using input-output models, the time-domain approach is difficult and frequency-domain or z-transform methods depend on stationarity of the system, which is not necessary for the state-space analysis of the operation of the system over a finite period of time. Also, nonlinear problems can be dealt with to some extent by linearization. Finally, and very importantly from a statistical point of view, if the noise-processes are Gaussian and white (by preliminary transformation or pre-whitening if necessary) and if the distribution of $\underline{x}(1)$ is Gaussian, then the posterior distribution of $\underline{x}(t)$, given $y(1),\ldots,y(t-1)$, is also Gaussian, with mean $\hat{\underline{x}}(t)$ and covariance matrix $P(t)$, say. What is more, these parameters can be computed recursively from

$$\hat{\underline{x}}(t+1) = G\hat{\underline{x}}(t) + Hu(t) + K(t)\nu(t) \tag{4}$$

$$y(t) = B\hat{\underline{x}}(t) + Cu(t) + \Sigma(t)\nu(t) \tag{5}$$

$$P(t+1) = GP(t)G - K(t)K^T(t) + FQF^T \tag{6}$$

where $K(t)$, called the gain, is defined by

$$K(t) = GP(t)B^T(BP(t)B^T + R)^{-1/2}\ , \tag{7}$$

and

$$\Sigma(t) = (BP(t)B^T + R)^{1/2}\ .$$

$$R = \mathrm{cov}\{n(t)\} \quad \text{and} \quad Q = \mathrm{cov}\{e(t)\}\ .$$

Often (4) and (5) are combined by elimination of the so-called innovation process $\nu(t)$ to give the updated $\hat{\underline{x}}(t+1)$ directly in terms of $\hat{\underline{x}}(t)$, $u(t)$ and $y(t)$. The innovation process can be shown to be a sequence of uncorrelated "standardized" normal random

variables or vectors; see Kailath (1968), Frost (1968).

Equations (4), (5) and (6) form the Kalman (-Bucy) filter (Kalman and Bucy, 1961), which is reappearing more and more in the statistical literature (Harrison and Stevens, 1976; O'Hagan, 1978). In the above they arise very naturally through Bayes Theorem and they have been derived in several other ways: as a recursive least squares algorithm, by maximum likelihood, using projection arguments, and by the so-called innovations approach. Stochastic approximation methods have also been used in the theory of the Kalman filter. The book by Jazwinskii (1970, Chapter 7) illustrates some of the approaches; see also Willems (1978).

If the system is stationary and the "steady-state" equilibrium has been reached, representations (4) and (5) have the interesting feature that, because the noise process $\{\nu(t)\}$ is the same in both, they can be replaced by an input-output model described by the z-transform equation

$$Y(z) = \{B(z^{-1}I-G)^{-1}H + C\}U(z) + \{B(z^{-1}I-G)^{-1}K + \Sigma\}\nu(z) \ . \qquad (8)$$

The <u>steady-state gain</u> K is computed by suppressing the time arguments in (6) and (7) and eliminating the resulting P.

Perhaps this is the point to mention that linear systems theory and Kalman filtering methods extend to more general situations. Continuous time processes are an obvious example, for which the equation governing the change of state will be a first order differential equation. An even more general set-up for which the Kalman theory carries over is that involving so-called <u>distributed-parameter</u> systems. Here the state is a function $\underline{x}(t,\omega)$ of two variables, where ω can take uncountably many values. In an application it

might represent spatial variables. In continuous time the model might take the form

$$\frac{\partial}{\partial t}\,\underline{x}(t,\omega) = L_{\omega}\,\underline{x}(t,\omega) + H(\omega)u(t,\omega) + \text{noise}\ ,$$

where L_{ω} is a linear operator involving ω-derivatives. In contrast, the models we have considered so far are so-called lumped-parameter systems. They correspond to ω taking only countably or finitely many values, so that a separate set of state variables can be defined for each ω-value. O'Hagan (1978) has used models similar to distributed-parameter systems; a bibliography of the field is available in Polis and Goodson (1976).

We return to the lumped-parameter case. The value of $\hat{\underline{x}}(s)$ is a natural point estimate of $\underline{x}(s)$, allowing fulfillment of the activity known as state-estimation or state-identification. If we have available the observations $y(1),\ldots,y(t-1)$, then, according as $s < t$, $s = t$, or $s > t$, the problem is one of smoothing, filtering, or prediction.

Although we shall briefly mention the problem of state-estimation, we shall concentrate more on that of parameter-estimation or system-identification. Our design problem is to select suitable inputs over a specified period of time, possibly infinite, to estimate the parameters "as well as possible"; optimal input signal synthesis. As in optimal experimental design, some criterion of efficiency will be proposed, and there generally will be some constraints on the allowable inputs. Our problem is different from that of optimal control. There the inputs must be chosen to keep the state vector as close as possible to some trajectory, or to home in on some target.

The extra dimension of time has major consequences as far as optimal design is concerned. We shall find that the attractive linear regression theory is rarely applicable, but we shall have the possibility of sequential design open to us. A nonsequential design, in which the input strategy is specified before the start, could be used, corresponding to off-line operation, but it will seem better to choose inputs as we go along, on-line, operating an adaptive system.

Two final concepts should be mentioned, controllability and observability, which are closely linked with the identifiability of parameters. Consider the deterministic model

$$\underline{x}(t+1) = G\underline{x}(t) + Hu(t)$$

$$y(t) = B\underline{x}(t) .$$

This system is completely observable if, after sufficient observations, the initial state $\underline{x}(1)$ can be exactly determined. It is completely controllable if, after sufficient stages, or choices of inputs, it is possible to translate the state to any specified position. These concepts are important if $u(\cdot)$ and $y(\cdot)$ are vectors, and elegant equivalent criteria exist in terms of the matrices G, H and B; see Wishart (1969).

From our point of view, the interesting point is that if system (3) is completely observable and controllable, then the parameters in the Kalman filter representation (4), (5), (6) are identifiable (Kailath, 1968, Appendix 2). Our estimation problems for the system therefore revolve around this model, or the equivalent input-output relationship (8); see, in particular, Section 4 D.

Before attacking the optimal design problem for dynamic systems, it is helpful to summarize the main results in "static" theory. Helpful textbooks on dynamic systems include Cadzow (1973), Astrom (1970), Eykhoff (1974), and Jazwinskii (1970).

3. OPTIMAL REGRESSION DESIGN

The static version of the problem concerns optimal design for regression models. Although O'Hagan (1978) uses Bayesian methods to develop quite a general approach, the usual starting point involves observations of a response function which depends on k unknown parameters θ, and on the site of the observation, u. The point u is chosen from some compact design space U. The problem is to decide how to distribute the available observations amongst the possible sites, or to choose an optimal design measure, which specifies the proportion of the observations to be made at the different points in U. The latter problem, which gives "approximate" designs, is theoretically much easier than the practically-motivated exact theory. The meaning of "optimal" will be discussed presently.

The basis of the theory (see the references in Section 1) was developed just for linear regression models with independent errors, but it is helpful to consider a more general set-up. Let us retain the feature of independent errors, but assume that the response function is nonlinear, in θ.

Let $I(\theta,u)$ denote the Fisher information matrix corresponding to an observation at u. Then, if a design ξ on U is used, the average per observation information matrix is

$$M(\theta,\xi) = \int_U I(\theta,u)\xi(du) \ .$$

Let ϕ be a real-valued convex decreasing function on the set of $k \times k$ nonnegative definite symmetric matrices, that is, such that $\phi(A) \leq \phi(B)$ if the matrix $A - B$ is nonnegative definite.

A <u>ϕ-optimal design</u> will be a design ξ^* such that $\phi(M(\theta,\xi))$ is <u>minimized</u> at ξ^*. Traditional choices for ϕ include

(i) $\phi(\cdot) = -\log\det(\cdot)$ (<u>D-optimality</u>)

(ii) $\phi(\cdot) = \mathrm{tr}(\cdot)^{-1}$ (<u>trace-optimality</u>) .

Let the Frechét directional derivative of ϕ from A in the direction of B be denoted by $\phi(A,B)$, and let $\underline{\Xi}$ denote the class of design measures on U.

Then we have the following:

<u>Theorem 1</u>.

For any θ,

(i) $M(\theta,\xi)$ is symmetric and nonnegative definite, for any $\xi \in \underline{\Xi}$;

(ii) $M(\theta) = \{M(\theta,\xi):\xi \in \underline{\Xi}\}$ is convex and compact;

(iii) the extreme points of $M(\theta)$ are each of the form $I(\theta,u)$, for some u; further, for any $\xi \in \underline{\Xi}$, there exists $n \in \underline{\Xi}$, assigning positive weight to at most $\frac{1}{2}k(k+1)+1$ points in U and such that $M(\theta,\xi) = M(\theta,n)$. (Essentially Carathéodory's Theorem.)

<u>Theorem 2</u>. (cf. Whittle, 1973; White, 1973)

For any θ, the following are equivalent:

(i) $\phi(M(\theta,\xi))$ is minimized at $M(\theta,\xi^*)$;

(ii) $\phi(M(\theta,\xi^*),M(\theta,n)) \geq 0$ for all $n \in \underline{\Xi}$.

If ϕ is differentiable at $M(\theta,\xi^*)$, we also have the equivalents:

(iii) $\phi(M(\theta,\xi^*), I(\theta,u)) \geq 0$ for all $u \in U$;

(iv) $\phi(M(\theta,\xi^*), I(\theta,u)) = 0$ for any u weighted positively in ξ^*, that is, for any u in the support of ξ^*.

(Thus Theorem 1 (iii) implies that a ϕ-optimal design can be achieved with finite support, and Theorem 2 (iv) gives a practical check for optimality when ϕ is differentiable.)

Algorithm 1.

Suppose $\phi(\cdot)$ is differentiable and that $\{\alpha_n\}$ is a sequence of numbers such that $0 < \alpha_n < 1$, $\alpha_n \to 0$ as $n \to \infty$, and $\sum_n \alpha_n = \infty$. Let u_n be the $u \in U$ that minimizes

$$\phi\{M(\theta,\xi_n), I(\theta,n)\} .$$

Then, from an initial $\xi_0 \in \Xi$ and subject to certain conditions, the sequence of designs generated by

$$\xi_{n+1} = (1 - \alpha_n)\xi_n + \alpha_n \xi(u_n)$$

converges to a ϕ-optimal design. $\xi(u_n)$ denotes the degenerate design concentrated on u_n.

Bound.

If $\phi(\cdot)$ is differentiable, $\xi \in \Xi$ and ξ^* is an optimal design, then

$$\phi(M(\theta,\xi)) - \phi(M(\theta,\xi^*)) \leq - \min_u \phi\{M(\theta,\xi), I(\theta,u)\} . \qquad (9)$$

This indicates how "close" ξ is to the optimum.

There are many versions of the equivalence theorem, Theorem 2, both in the statistical and in the engineering literature, but the proofs are almost all essentially the same as that of Whittle (1973), which, because of its generality, is satisfyingly simple and elegant. The only radically different approach is that described for D-optimality in terms of Lagrangian duality theory by Sibson (1972, 1974) and Silvey and Titterington (1973).

Algorithm 1 is only one of many that have been suggested for the computation of optimal designs; see Fedorov (1972), Wu and Wynn (1978), Wu (1976), St. John and Draper (1975) and Titterington (1977). There is a fundamental snag to its application, and that is its dependence on θ. It seems that we have to know the true value of θ in order to calculate an optimal design for estimating it! In the special case of linear regression ("linear" in θ) it is easy to see that the θ-dependence disappears, so that an optimal design can, in principle, be computed before the experiment starts. In this case $M(\cdot)$ is proportional to the inverse of the covariance matrix of the least-squares estimator of θ. In the nonlinear case there are three possible general approaches.

(i) Apply Algorithm 1 with a prior estimate, θ_0, of θ, and generate an "off-line" design as for the linear case.

(ii) Propose some weighting function $W(\cdot)$ on the parameter space, Ω. This may or may not be a formal prior density. Then construct either

$$M(\xi) = \int_{\Omega} M(\theta,\xi)W(d\theta)$$

or a new criterion

$$\phi_W(\xi) = \int_\Omega \phi\{M(\theta,\xi)\}W(d\theta) \ .$$

In both cases equivalence theorems can be written down and proved on the usual way; see Mehra (1974b), and Läuter (1974).

(iii) Carry out a sequential design procedure, or on-line input synthesis. One such modification is the following:

Algorithm 2.

Suppose ϕ is differentiable and that after n observations have been made, corresponding to a design ξ_n, an estimate $\hat{\theta}_n$ is available for θ.

Suppose u_n minimizes $\phi\{M(\hat{\theta}_n,\xi_n), I(\hat{\theta}_n,u)\}$.

Take an observation at u_n, set $n = n + 1$, update the design measure and repeat the procedure.

Algorithms of this type have been considered by Fedorov and Malyutov (1972), White (1975), and Ford (1976), and in the engineering literature, as will be reported later. Chernoff (1953) discussed the awkward dependence of optimal designs on θ and he encouraged the development of sequential procedures in Chernoff (1975).

Convergence of these algorithms is very hard to prove. It would be plausible if the sequence $\{\hat{\theta}_n\}$ were consistent, but this itself is difficult because of the complicated statistical properties of, say, the sequence of maximum likelihood estimators of θ; see White (1975), Ford (1976), and Goodwin and Payne (1977, p. 115). Algorithm 2 does,

however, seem to be a workable and helpful way of dealing with the problem of nonlinearity. Admittedly, it is only a "one-step-ahead" procedure, but more sophisticated methods incur very heavy computations. This is evident in the Bayesian approach of O'Hagan (1978), and is one feature that merits further research.

This concludes a brief summary of the major results in optimal experimental design. We only mention the important topics of exact designs, designs for special purposes such as model choice, and the problems that arise with designs with singular information matrices, which may lead to lack of differentiability of ϕ. This occurs particularly when only a subset of the parameters is of interest.

4. EXTENSION OF OPTIMAL DESIGN THEORY TO INPUT SIGNAL DESIGN IN DYNAMIC SYSTEMS.

In this section we illustrate how the results from Section 3 find application in the input-synthesis problem in dynamic system identification, as the engineers would describe it. A few typical papers are summarized and reference is made to other similar work.

A. Box and Jenkins (1976, Appendix A 11.2)

The model considered here is a very simple input-output relationship, but it brings out several important points.

$$y(t+1) - a_1 y(t) = b_1 u(t) + e(t) ,$$

where all variables are scalar, $|a_1| < 1$ and $\{e(t)\}$ is white Gaussian noise. (In the following, the engineering terminology will be introduced more and more.) The parameters a_1 and b_1 are to be estimated.

Box and Jenkins consider choosing input processes to maximize the determinant of the long-term information matrix, subject to con-

straints on the input. They consider constraints of the form

(i) σ_u^2 fixed, (ii) σ_y^2 fixed, (iii) $\sigma_u^2\sigma_y^2$ fixed.

The solutions are that in cases (i) and (ii) first-order autoregressions are optimal, whereas in case (iii) white noise is best. Major problems are that the parameters in the autoregressions depend on a_1 and b_1. In the dynamic case, therefore, linearity of the model in the parameters does not usually guarantee the possibility of off-line design.

Similar problems are discussed by Ng et al. (1977) for higher-order autoregressions. As shown by Levin (1960) the optimal input can be computed "off-line" if the input-output relationship is a moving average.

B. Zarrop, Payne and Goodwin (1975)

A more complicated stationary input-output representation is considered in this paper:

$$\sum_{i=0}^{n} a_i y(t-i) = \sum_{i=0}^{n} b_i u(t-i) + \sum_{i=0}^{n} c_i e(t-i) \ . \qquad (10)$$

Each of $y(t)$, $u(t)$ and $e(t)$ is considered to be a vector (multi-input-multi-output), the $\{e(t)\}$ are taken to be independent normal $(0,\Sigma)$, a_o and c_o are identity matrices and the parameters of interest, θ, are the elements of $(a_1,\ldots,a_n, b_o,\ldots,b_n, c_1,\ldots,c_n,\Sigma)$. From the log-likelihood of an N-sample the Fisher information matrix is computed. The asymptotic per observation information matrix is expressed in terms of the normalized spectral distribution function, F_u, of the input process and it can be written as

$$M(\theta,F_u) = \text{constant} + \text{term "linear" in } dF_u(\cdot) \ .$$

The input spectral density takes the role of the design measure in Section 3 and theorems exactly analogous to Theorems 1 and 2 are available, although the proof of the latter is not given in this reference, but in Goodwin and Payne (1977) and Mehra (1974b, 1976a). An important modification to Theorem 1 is the definition of the extreme points of the design space, which is now the range of frequencies $(-\pi,\pi)$. The extreme points correspond to pure sine-wave inputs and Carathéodory's Theorem implies that, if θ is k-dimensional, an optimal input process can be constructed as a linear combination of at most $\frac{1}{2}k(k+1)$ sinusoids. This frequency-domain approach has a startling advantage in that the design-space is a finite interval, although again the specific optimal frequencies and amplitudes are θ-dependent. Mehra (1976a) suggests substituting a prior estimate θ_0, and proposes a version of Algorithm 1 for computing the optimal design on $(-\pi,\pi)$.

Other papers related directly to this problem are Payne et al. (1975) and Viort (1972). The latter work seems to have been the first attempt to investigate D-optimality in dynamic systems. In general, all these papers concentrate on D-optimality, although the basic theorems have much wider validity, as in Section 3.

C. Keviczky (1975)

Keviczky considers a scalar (single-input-single output) version of (10), specifically of the form

$$\sum_{i=0}^{n} a_i y(t-i) = \sum_{i=0}^{m} b_i u(t-i) + \lambda \sum_{i=0}^{r} c_i e(t-i)$$

where $a_0 = c_0 = 1$, $m \leq n$, $r \leq n$. The errors are identically distributed and Gaussian, with zero means. He considers separately the

case of $\{e(t)\}$ uncorrelated and correlated and he works with finitely many observations and therefore in the time-domain. As in B, he constructs the Fisher information matrix, regarding as the k parameters $\theta^T = (b_0,\ldots,b_m, a_1,\ldots,a_n)$, and, for the uncorrelated errors case, derives a recursion for the determinant of the covariance matrix of the least-squares estimators of the parameters, from the N-observation case to that of $N + 1$. Using this, he is able to choose $u(N + 1)$ to maximize the increase in the D-optimality criterion subject to an amplitude constraint

$$-U \leq u(N + 1) \leq U .$$

As often happens, it is optimal to take $|u(N + 1)| = U$.

When the errors are correlated, optimal design has to be based on the information matrix, which involves the usual difficulty of ignorance about θ.

A summary of this approach is given by Goodwin and Payne (1977), and recursive design is also described by Arimoto and Kimura (1971), using a Bayesian information-theoretic viewpoint.

D. Mehra (1974b)

This paper reviews the field quite fully and discusses explicitly state-space models like (3). Controllability and observability are assumed and the identifiable form of the system given by (4) - (6) is considered. As in B and C, the log-likelihood from N observations is written down as

$$L(\theta) = \text{constant} - \frac{1}{2} \sum_{t=1}^{N} \left\{ \nu^T(t)\nu(t) + 2|\Sigma(t)| \right\}$$

where θ denotes the set of k parameters. The Fisher information matrix related to, say, a scalar input sequence $u_N^T = (u(1),\ldots,u(N))$

can be evaluated, after some effort (details in Mehra, 1976b), and it turns out that

$$M(\theta,\xi) = \int_{u_N} W(\theta,\xi(u_N))\xi(du_N) + A(\theta) = W(\theta,\xi) + A(\theta)$$

where ξ is a measure on the space u_N of all possible u_N, and each element of $W(\theta,u_N)$ is a quadratic form in u_N. Typical input constraints, defining u_N, are

(a) $u_N^T u_N \leq 1$ (energy constraint)

(b) $|u(t)| \leq 1$, $t = 1,\ldots,N$ (amplitude constraints).

Again, D-optimality is considered and direct analogues of Theorems 1 and 2, Algorithm 1 and the Bound (Mehra, 1976a) are provided.

In this special case,

$$\phi\{M(\theta,\xi),I(\theta,u_N)\} = \mathrm{tr}\{M^{-1}(\theta,\xi)W(\theta,\xi)\} - \mathrm{tr}\{M^{-1}(\theta,\xi)W(\theta,\xi(u_N))\},$$

and the equivalence theorem is given in terms of this. The formulation given here follows Mehra (1976a, pp.230-249) more closely. In Mehra (1974b) a prior distribution is assumed for θ and results given in terms of

$$M(\xi) = E_\theta M(\theta,\xi).$$

Implementation of Algorithm 1 involves, at stage n, the choice of a $u_N^{(n)} \in u_N$ to maximize

$$\mathrm{tr}\{M^{-1}(\theta,\xi_n)W(\theta,\xi(u_N))\}, \qquad (11)$$

and this itself is a complex procedure. Criterion (11) is quadratic in u_N, so in the bounded-energy case (a) we have an eigenfunction problem and in the linearly-constrained bounded-amplitude case (b) a quadratic programming problem, which results in a bang-bang input;

that is, $|u(t)| = 1$, for each t.

Even when an optimal design on u_N has been computed, it is not clear how to apply it, being a probability measure on the set of N-stage inputs. Mehra (1974b) suggests some "concatenation" in which parts of the positively weighted $\{u_N\}$ (those in the support of the design) are applied in turn.

In the stationary case, Mehra considers computation of long-term optimal systems by frequency-domain analysis as in B above, pointing out the computational attractiveness related to the simple design space $(-\pi,\pi)$. This approach is made possible by the possibility of representing the equilibrium model in input-output terms as in (8) in Section 2.

Mehra (1974b) indicates generalizations to continuous-time, nonlinear and distributed parameter systems and gives some continuous-time illustrative examples. The only complication in the continuous-time frequency-domain approach is that the design space of frequencies is $(-\infty,\infty)$.

E. Aoki and Staley (1970)

Although this paper is more tenuously dependent on the work of Section 3, it is appropriate to mention it now because it uses an optimality criterion related to B above. The authors consider the autoregressive model

$$\sum_{i=0}^{k} a_i y(t-i) = u(t) + e(t) , \qquad t = 1,2,...$$

where the $\{e(t)\}$ are independent $N(0,\sigma^2)$, $a_0 = 1$ and suitable initial conditions are specified. N observations have to be made to estimate $\theta^T = (a_1,...,a_k)$ and the criterion used is

$$-\text{tr } M(\theta,\xi) \qquad (12)$$

where ξ is a design on u_N and M is the Fisher information matrix.

For this criterion, we can find an optimal design that is degenerate, concentrated on one point in the design space, that is, on one set of N inputs. In the bounded energy case, therefore, we must maximize

$$\text{tr}(I(\theta,u_N)) ,$$

which turns out to be quadratic in u_N, subject to the quadratic constraint $u_N^T u_N \leq 1$. As in the treatment of (11) in D above, we must solve an eigenproblem.

The continuous-time version of this problem expressed in state-space terms has been examined by Mehra (1974a). Instead of N discrete inputs, observations are made over a finite period $(0,T)$. The trace criterion (12) is used, and the energy constraint is

$$\int_0^T u(t)^T u(t)dt = 1 .$$

The optimal $\{u(t): 0 \leq t \leq T\}$ can again be regarded as an eigenfunction. When there is a scalar parameter the problem is of Sturm-Liouville type and the equation satisfied by the optimal input can be regarded as a Fredholm integral equation. Various methods are suggested for obtaining explicit solutions.

The trace criterion (12) (not the same as the usual one of $\text{tr}(\cdot)^{-1}$) was also used by Nahi and Napjus (1971) and Lopez-Toledo (1974). Its use was criticized by Zarrop and Goodwin (1975), with rejoinder by Mehra (1975), on the grounds that the information matrices corresponding to the optimal designs are often singular, so that identifiability problems may well arise. The corresponding trivial

observation in the static design case is implicit in Silvey and Titterington (1974).

F. Dogorovcev (1971), and Spruill and Studden (1978)

In these papers the influence of time is felt in a slightly different way. In Dogorovcev the response function is

$$y(t) = \theta^T u(t) + e(t) , \quad 0 \le t \le T \tag{13}$$

where θ and $u(t)$, for each t, are k-vectors and $\{e(t)\}$ is a stationary process with zero means and covariance kernel $R(s,t)$, defining a reproducing kernel Hilbert space (RKHS) $H(R)$, in the sense of Parzen (1961). θ is estimated by $\hat{\theta}$, the best estimator linear in $\{y(t): 0 \le t \le 1\}$ and optimal functions $\{u(t): 0 \le t \le 1\}$ are sought for trace optimality of $\mathrm{cov}(\hat{\theta})$ and to minimize the variance of $c^T\hat{\theta}$, for a specified vector c. An orthogonal basis can be set up in $H(R)$ and approximately normalized members of this basis provide the optimal input functions.

Spruill and Studden (1978) extend the work to more complex responses, dependent on a spatial variable ω as well as t.

G. Miscellaneous papers

In Mehra (1974a) it was not necessary to choose a design on the class of input strategies because a degenerate design was optimal. Such a search for an optimal input process leads to a more conventional numerical problem. The choice of inputs over the interval $(0,T)$, subject to some integral constraint, to minimize some time-integrated functional is a familiar extremal problem in control engineering, leading to solution by variational methods or by the theory of eigenfunctions. Such methods constitute what is sometimes

called the optimal control approach.

Mehra (1974a) falls into this category, as does his precursor Levadi (1966). This author considered the continuous-time scalar moving-average process

$$y(t) = \int_0^T b(t,\tau)u(\tau)d\tau + e(t) , \qquad 0 \le t \le T ,$$

where the function $b(\cdot,\cdot)$ contains k parameters, θ. An input process $\{u(t): 0 \le t \le T\}$ has to be chosen, subject to $\int_0^T u^2(t)dt = 1$, to minimize $\operatorname{tr} I^{-1}(\theta,u(\cdot))$, where $I^{-1}(\cdot,\cdot)$ denotes the covariance matrix of the best linear estimator of θ. The noise may be correlated. In this moving-average example the optimal input is independent of θ, unlike most of the cases we have considered, and the solution, which, like Dogorovcev (1971), is based on the RKHS formulation of Parzen (1961), is again an eigenfunction satisfying a Fredholm integral equation.

In Goodwin (1971) a discrete-time nonlinear state-space system, involving k parameters θ, coupled with a linear observation equation, is linearized. The performance-index to be minimized is the sum of the more usual trace criterion (see Section 3) and a penalty function to restrict the input choice. Numerical solution is necessary; see also Nahi and Wallis (1971).

The trace criterion has been used in single-input-single-output input-output models by Goodwin, et al. (1973), Goodwin and Payne (1973). Hamiltonian methods were used to compute optimal inputs which compared well with suboptimal strategies.

Other approaches have considered specific types of input process and tried to optimize within the appropriate class; see Van der Bos (1967, 1973), Litman and Huggins (1963).

5. OPTIMAL CHOICE OF SAMPLING INSTANTS OR LOCATIONS

Section 4 concentrated on the problem of choosing optimal input processes. Here we consider how, given the input processes for a continuous-time system that can be sampled at discrete instants, the sampling strategy should be designed. Three models are considered below, corresponding to three groups of authors.

A. Sacks, Ylvisaker and Wahba

These authors considered the response models similar to (13) in Section 4F. Sacks and Ylvisaker (1966) looked at the scalar parameter case, with

$$y(t) = \theta u(t) + e(t) , \quad 0 \leq t \leq 1 ,$$

and they considered how to choose N distinct sampling times so as to minimize the variance of the best linear estimator of θ. In particular, they looked at what happened as $N \to \infty$ and found that asymptotically optimal solutions existed, provided the error covariance kernel $R(s,t)$ was non-smooth on $s = t$ and provided $u(t)$ belonged to $H(R)$. The optimal solutions were characterized but their explicit computation is difficult. In later papers (1968, 1970), they extended their work to the k-parameter case and relaxed the aforementioned condition on $R(s,t)$.

Wahba (1971) related the problem to one of function approximation by splines and did more work on the computation of optimal sequences of designs.

B. Goodwin, Mehra and others

In a series of papers starting with Goodwin, et al. (1974), linear state-space models of the following form were considered.

$$d\underline{x}(t) = G\underline{x}(t)dt + Hu(t)dt + Fde(t) \qquad (14)$$
$$y(t) = B\underline{x}(t) + Cu(t) \quad ,$$

with some initial conditions, where θ, the parameter vector, is made up of the unknowns in (G, H, B, C) and where $\{e(t)\}$ is a (multivariate) process with independent increments. Sampling times for N observations have to be chosen, which leads to the replacement of (14) by an appropriately integrated discrete recursion. The Fisher information matrix is written down and the D-optimality and trace-optimality criteria are considered. In Goodwin, et al. (1974) and Payne, et al. (1975), where for N large the frequency-domain approach is used, the improvement resulting from non-uniform sampling intervals in a simple example and the simultaneous choice of sampling frequency and input function, are discussed. The work is summarized in Goodwin and Payne (1976, 1977, Section 6.5).

Mehra (1976b) considers a similar model to (14), with observation equation

$$y(t) = B\underline{x}(t) + n(t) \quad , \qquad 0 \leq t \leq 1$$

where $n(t)$ is Gaussian white noise. He considers the choice of measurement times not now from a parameter estimation point of view but from one of state estimation. Instead of the Fisher information matrix he considers $P(t)$, the covariance matrix of the state vector, whose evolution is governed by the Kalman equations. In particular, he would like to "minimize" $P(1)$. However, $P(1)$ is hard to compute explicitly in the continuous-time model, as a solution of a Riccati equation and he opts for approximations that lead back to consideration of criteria based on the Fisher information matrix. The paper closely parallels Mehra (1974b), with results like Theorems 1 and 2

and a computational procedure on the lines of Algorithm 1.

Ng and Goodwin (1976) consider the model

$$d\underline{x}(t) = G\underline{x}(t)dt + Fde(t)$$

$$y(t) = B\underline{x}(t) \quad ,$$

where $e(t)$ is white Gaussian noise.

Using $(0,\infty)$ as the design space, they represent a sampling strategy as a design ξ on $(0,\infty)$ with, as extreme points, the uniform sampling rates. The appropriate Fisher information matrix satisfies Theorem 1, so that in a k-parameter problem an optimal strategy can be achieved using at most $\frac{1}{2}k(k+1)$ sampling rates.

As usual, explicit results depend on knowledge of θ. Methods of coping with ignorance about θ and of frequency-domain analysis are described. They show that it is better to concatenate subexperiments using pure sinusoids than to carry out a single experiment with a multi-frequency input.

C. Seinfeld and others

These authors are involved with such spatial problems as the measurement of pollution and with the location of monitoring stations. They are therefore obliged to look at distributed-parameter systems and to construct spatial designs that are both exact (a general design measure will not do) and non-replicating.

With the pollution levels as state variables, the Kalman filter equations are constructed and a criterion of "total integrated variance" (integrated over time and space) is computed, using the Riccati equation for the state covariance matrix $P(t,\omega)$. Heavy numerical work is necessary for optimal choice of, say, N locations.

In Yu and Seinfeld (1973) sequential choice is suggested; see also Chen and Seinfeld (1975), Seinfeld (1976).

6. CLOSING REMARKS

The problem of optimal experimental design is clearly important in models where time is an inevitable component. It is pleasing that the Kiefer-Wolfowitz work in static models carries over automatically to many dynamic problems in which the underlying design space can be a class of input processes (time-domain analysis), a range of frequencies (frequency-domain) or a range of sampling rates (sampling strategies).

Although the published theory does not go far beyond D-optimality or trace-optimality, general criteria could be considered, as in Section 3. There is not much written about D_s-optimality and its counterparts, in which only s of the k parameters are of interest, although again this would not involve any extra fundamental ideas. Goodwin and Payne (1977) mention it in the context of model discrimination, drawing analogy with work of Atkinson and Cox (1974) and Atkinson and Fedorov (1975a, 1975b). The latter two papers bear some resemblance to a paper by Gagliardi (1967), who considers a special model choice problem involving a finite parameter space. His methods are generalized by Kuszta and Sinha (1976, 1977, 1978).

These are some areas where further development is desirable. Another is the consideration of more examples, in particular less simple ones than at present reported, although they are likely to involve hard computational work because of the almost inevitable dependence of the information matrix on the parameter values. Any advances in the numerical aspects of these problems would be very valuable.

REFERENCES

Aoki, M. and Staley, R.M. (1970). On input signal synthesis in parameter identifications. Automatica, 6, 431-440.

Arimoto, S. and Kimura, H. (1971). Optimum input test signals for system identification - an information-theoretical approach. Int. J. Syst. Sci., 1, 279-290.

Astrom, K.J. (1970). Introduction to Stochastic Control Theory. New York: Academic Press.

Atkinson, A.C. and Cox, D.R. (1974). Planning experiments for discriminating between models. J. R. Statist. Soc. B, 36, 321-348.

Atkinson, A.C. and Fedorov, V.V. (1975). The design of experiments for discriminating between two rival models. Biometrika, 62, 57-69.

_____ (1975). Optimal design: experiments for discriminating between several models. Biometrika, 62, 289-303.

Box, G.E.P. and Jenkins, G.M. (1976). Time Series Analysis Forecasting and Control. San Francisco: Holden-Day.

Cadzow, J.A. (1973). Discrete-Time Systems. Englewood Cliffs: Prentice-Hall.

Chen, W.H. and Seinfeld, J.H. (1975). Optimal location of process measurements. Int. J. Control, 21, 1003-1014.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. Ann. Math. Stat., 24, 586-602.

_____ (1975). Approaches in sequential design of experiments. In Statistical Design and Linear Models, Ed. J. N. Srivastava, 67-90. Amsterdam: North-Holland.

Dogorovcev, A.J. (1971). Problems of optimal control of a regression experiment. Selected Transl. in Math. Stat. and Prob., 10, 35-41. AMS.

Eykhoff, P. (1974). System Parameter and State Estimation. London: Wiley.

Fedorov, V.V. (1972). Theory of Optimal Experiments. New York: Academic Press.

Fedorov, V.V. and Malyutov, M.B. (1972). Optimal designs in regression problems. Math. Operationsforsch. u. Stat., 3, 281-308.

Ford, I. (1976). Ph.D. Thesis. University of Glasgow.

Frost, P.A. (1968). Ph.D. Thesis. Stanford University.

Gagliardi, R.M. (1967). Input selection for parameter estimation in discrete systems. IEEE Trans. Automat. Contr., AC-12, 597-599.

Goodwin, G.C. (1971). Optimal input signals for nonlinear system identification. Proc. Inst. Elec. Eng., 118, 922-926.

Goodwin, G.C., Murdoch, J.C. and Payne, R.L. (1973). Optimal test signal design for linear single input-single output system identification. Int. J. Control, 17, 45-55.

Goodwin, G.C. and Payne, R.L. (1973). Design and characterization of optimal test signals for linear single input-single output parameter estimation. Proc. IFAC Symp. Identification, The Hague/Delft, 1973, 1005-1010.

_____ (1976). Choice of sampling intervals. In System Identification: Advances and Case Studies, Eds. R. K. Mehra and D.G. Lainiotis, 251-287. New York: Academic Press.

_____ (1977). Dynamic System Identification: Experiment Design and Data Analysis. New York: Academic Press.

Goodwin, G.C., Zarrop, M. and Payne, R.L. (1974). Coupled design of test signals, sampling intervals and filters for system identification. IEEE Trans. Automat. Contr., AC-19, 748-752.

Harrison, P.J. and Stevens, C.F. (1976). Bayesian forecasting. J. R. Stat. Soc.B, 38, 205-247.

Jazwinskii, A.H. (1970). Stochastic Processes and Filtering Theory. New York: Academic Press.

Kailath, T. (1968). An innovations approach to least-squares estimation - part 1: Linear filtering in additive white noise. IEEE Trans. Automat. Contr., AC-13, 646-655.

Kalman, R.E. and Bucy, R.S. (1961). New results in linear filtering and prediction theory. J. Basic Eng. D, 83, 95-107.

Keviczky, L. (1975). Design of experiments for the identification of linear dynamic systems. Technometrics, 17, 303-308.

Kiefer, J. (1959). Optimum experimental designs. J. R. Stat. Soc. B, 21, 273-319.

_____ (1974). General equivalence theory for optimum designs (approximate theory). Ann. Stat., 2, 849-879.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. Can. J. Math., 12, 363-366.

Kuszta, B. and Sinha, N.K. (1976). Optimum input signals for parameter estimation. Int. J. Syst. Sci., 7, 935-941.

_____ (1977). Int. J. Syst. Sci., 8, 1097.

_____ (1978). Design of optimal input signals for the identification of distributed parameter systems. Int. J. Syst. Sci., 9, 1-7.

Lauter, E. (1974). Experimental planning in a class of models. Math. Operationsforsch. u. Stat., 5, 323-341.

Levadi, V.S. (1966). Design of input systems for parameter estimation. IEEE Trans. Automat. Contr., AC-11, 205-211.

Levin, M. J. (1960). Optimal estimation of input response in the presence of noise. IRE Trans. Circuit Theory, CT-7, 50-56.

Litman, S. and Huggins, W.H. (1963). Growing exponentials as a probing signal for system identification. Proc. IEEE, 51, 917-923.

Lopez-Toledo, A.A. (1974). Ph.D. Thesis. M.I.T.

Mehra, R.K. (1974a). Optimal inputs for linear system identification. IEEE Trans. Automat. Contr., AC-19, 192-200.

_____ (1974b). Optimal input signals for parameter estimation in dynamic systems - survey and new results. IEEE Trans. Automat. Contr., AC-19, 753-768.

_____ (1975). Rejoinder to Zarrop and Goodwin (1975). IEEE Trans. Automat. Contr., AC-20, 300.

_____ (1976a). Synthesis of optimal inputs for multi-input - multi-output systems with process noise. In System Identification: Advances and Case Studies, eds. R.K. Mehra and D. G. Lainiotis, 211-250. New York: Academic Press.

_____ (1976b). Optimization of measurement schedules and sensor designs for linear dynamic systems. IEEE Trans. Automat. Contr., AC-21, 55-64.

Nahi, N.E. and Napjus, G.A. (1971). Design of optimal probing signals for vector parameter estimation. Proc. IEEE Conf. on Decision and Control, Miami.

Nahi, N. E. and Wallis, D.E. (1969). Optimal inputs for parameter estimation in dynamic systems with white observation noise. Proc. JACC, Boulder.

Ng, T. S. and Goodwin, G. C. (1976). On optimal choice of sampling strategies for linear system identification. Int. J. Control, 23, 459-475.

Ng, T. S., Goodwin, G. C. and Payne, R. L. (1977). On maximal accuracy estimation with output power constraints. IEEE Trans. Automat. Contr., AC-22, 133-134

O'Hagan, A. (1978). Curve fitting and experimental design for prediction. J. R. Statist. Soc. B, 40, 1-42.

Parzen, E. (1961). An approach to time series analysis. Ann. Math. Statist., 32, 951-989.

Payne, R. L., Goodwin, G. C. and Zarrop, M. (1975). Frequency domain approach for designing sampling rates for system identification. Automatica, 11, 189-191.

Polis, M. P. and Goodson, R. E. (1976). Parameter identification in distributed systems: a synthesizing overview. Proc. IEEE, 64, 45-61.

St. John, R. C. and Draper, N. R. (1975). D-optimality for regression designs: a review. Technometrics, 17, 15-23

Sacks, J. and Ylvisaker, D. (1966). Designs for regression problems with correlated errors. Ann.Math. Statist., 37, 66-89.

_____ (1968). Designs for regression problems with correlated errors: many parameters. Ann. Math. Statist., 39, 45-69.

_____ (1970). Designs for regression problems with correlated errors, III. Ann. Math. Statist., 41, 2057-2074.

Sibson, R. (1972). Contribution to discussion of papers by Wynn and Laycock, J. R. Statist. Soc. B, 34, 181-183.

Seinfeld, J. H. (1976). Optimal design for a pollutant monitoring system for an airshed. Proc. 1976 CISS, Baltimore, Md., 320-328. Dept. of Elec. Eng., Johns Hopkins Univ.

Sibson, R., (1974). D_A-optimality and duality. In Progress in Statistics, Proc. EMS, Budapest, 1972. Eds. J. Gani, K. Sarkadi, I. Vincze. Amsterdam: North Holland.

Silvey, S. D. (1974). Some aspects of the theory of optimal linear regression design with a general concave criterion function. Tech. Rep. 75, Series 2. Dept. of Stats., Princeton University.

Silvey, S. D. and Titterington, D. M. (1973). A geometric approach to optimal design theory. Biometrika, 60, 21-32.

_____ (1974). A Lagrangian approach to optimal design. Biometrika, 61, 299-302.

Spruill, C. and Studden, W. J. (1978). Optimum designs when the observations are second order processes. J. Multivar. Anal., 8, 153-172.

Titterington, D. M. (1977). A survey of optimal design algorithms. Proc. 10th EMS, Leuven, Belgium, 64-65.

Van Der Bos, A. (1976). Construction of binary multifrequency test signals. IFAC Symp. Identification, Prague, 1967.

_____ (1973). Selection of periodic test signals for estimation of linear system dynamics. IFAC Symp. Identification, The Hague/Delft.

Viort, B. (1972). D-optimal designs for dynamic models: Part I, Theory. Tech. Report 314, Dept. Statist., University of Wisconsin, Madison.

Wahba, G. (1971). On the regression design problem of Sacks and Ylvisaker. Ann. Math. Statist., 42, 1035-1053.

White, L. V. (1973). An extension of the General Equivalence Theorem to nonlinear models. Biometrika, 60, 345-348.

_____ (1975). Ph.D. Thesis. Imperial College, London.

Whittle, P. (1973). Some general points in the theory of optimal experimental design. J. R. Statist. Soc. B, 35, 123-130.

Willems, J. C. (1978). Recursive filtering. Statist. Neerlandica, 32, 1-39.

Wishart, D. M. G. (1969). A survey of control theory. J. R. Statist. Soc. A, 132, 320-334.

Wu, C-F. (1976). Ph.D. Thesis, University of California, Berkeley.

Wynn, H. P. (1974). Contribution to discussion of paper by Atkinson and Cox. J. R. Statist. Soc. B, 36, 335-336.

Wu, C.-F. and Wynn, H. P. (1978). General step-length algorithms for non-singular optimum design criteria. Ann. Statist. (to appear).

Young, P. C. (1974). Contribution to discussion of paper by Brown, Durbin and Evans. J. R. Statist. Soc. B, 37, 168-174.

Yu, T. K. and Seinfeld, J. H. (1973). Observability and optimal measurement location in linear distributed parameter systems. Int. J. Control, 18, 785-799.

Zarrop, M. and Goodwin, G. C. (1975). Comments on Mehra (1974a). IEEE Trans. Automat. Contr., AC-20, 299-300.

Zarrop, M., Payne, R. L. and Goodwin, G. C. (1975). Experiment design for time series analysis: the multivariate case. Tech. Rep. EE7506, Dept. Elec. Eng., University Newcastle, NSW, Australia.